# One Context Unification Problems Solvable in Polynomial Time

# One Context Unification Problems Solvable in Polynomial Time

Adrià Gascón and Ashish Tiwari
SRI International, Menlo Park, USA

Manfred Schmidt-Schauss
Goethe-Universität, Frankfurt, Germany

***Abstract*—One context unification extends first-order unification by introducing a single context variable, possibly with multiple occurrences. One context unification is known to be in NP, but it is not known to be solvable in polynomial time. In this paper, we present a polynomial time algorithm for certain interesting classes of the one context unification problem. Our algorithm is presented as an inference system that nontrivially extends the usual inference rules for first-order unification. The algorithm is of independent value as it can be used, with slight modifications, to solve other problems, such as the first-order unification problem that tolerates one clash.**

## I. Introduction

The problem of checking satisfiability of a set of equations plays a central role in any mathematical science. From the perspective of computer science, a lot of effort is devoted to finding efficient decision procedures for different families of equations. The problem of *satisfiability of word equations*, also known as *word unification*, figures prominently as one of the most intriguing problems of that form. The first algorithm for that problem was given by Makanin [15], and the best known upper bound (PSPACE) is due to Plandowski [19]. On the other hand, its PSPACE-hardness is an open question. Several particular cases of that problem, such as the ones that result from fixing the number of variables in the equations to a constant, have also been studied. For instance, efficient algorithms for satisfiability of word equations with one [4], [17], [11] and two [3] variables have been discovered.

Another fundamental operation in symbolic computation systems is the well-known *first-order unification problem*. This problem consists of solving equations of the form $s \doteq t$, where $s$ and $t$ are terms with first-order variables. The goal is to find a mapping from variables to (first-order) terms that would make the terms $s$ and $t$ syntactically equal. This problem was firstly introduced as such in the work by J.A. Robinson, which established the foundations of automated theorem proving and logic programming. More concretely, Robinson presented in [20] a procedure to determine the validity of a first-order sentence that has term unification as its main ingredient. Later, term unification was also used by Knuth and Bendix as a key component of their critical pairs method to determine local confluence of term rewrite systems (see [1] for a general survey on unification theory). The syntactic unification and matching problems were deeply investigated in the last century. Among other results, linear time algorithms were discovered [16], [18]. Moreover, more expressive variants of term unification such as *unification modulo theories* have also drawn a lot of attention. In this notion of term unification, equality between terms is interpreted under equational theories such as associativity, commutativity, and distributivity, among others [1].

An interesting connection between word and term unification is the *context unification problem*. In context unification, the terms $s, t$ in the equation $s \doteq t$ may contain context variables. For example, consider the equation $F(f(x, b)) \doteq f(a, F(y))$. where $x, y$ are first-order variables ranging over terms and $F$ is a context variable that can be replaced by any context. One of the possible solutions of this instance is the substitution $\{F \mapsto f(a, \bullet), x \mapsto a, y \mapsto b\}$. Note that when we instantiate $F$ by $f(a, \bullet)$ in the equation, replacing the occurrence of $\bullet$ by the argument of $F$ in each of its occurrences, we get $f(a, f(x, b)) \doteq f(a, f(a, y))$, and thus both sides of the equations become equal after applying $\{x \mapsto a, y \mapsto b\}$. Note that, simply using a unary signature, word unification reduces to context unification. On the other hand, context unification is a particular case of second-order unification, which is undecidable [9]. The decidability of context unification remained open for a long time, until recently a PSPACE algorithm was presented by Jeż [12].

Several variants of context unification with decision procedures in NP, such as stratified context unification and well-nested context unification, have been considered. Such variants have applications in computational linguistics and unification up to several positions [13], [5], [14]. Furthermore the variant of context unification where one of the sides of the equation is ground, called *context matching*, has also been investigated [21], [6] and, although the problem is in general NP-hard, polynomial time algorithms are known for some cases. In this paper we revisit the particular case of context unification where only one context variable, possibly with many occurrences, occurs in the input equations. This problem is known as *one context unification* (1-CU). In [7], a non-deterministic polynomial time algorithm for 1-CU was presented. That result has later been extended [2] also to the case where the input terms are represented with Singleton Tree Grammars, a grammar-based compression mechanism more general than Directed Acyclic Graphs. On the other hand, 1-CU is not known to be NP-hard nor solvable in polynomial time. This gap motivates the work described in this paper. We also remark here that initial interest in one context unification came from interprocedural program analysis [10], where context variables are used to represent (the yet unknown) summaries of procedures. In particular, one context unification problems (over uninterpreted terms) arise

when analyzing programs using an abstract domain consisting of (uninterpreted) terms.

### A. Related Work

A non-deterministic polynomial time procedure for one context unification was presented in [7]. There are instances where that algorithm provably takes exponential worst-case running time. For example, let $s$ be $f(x_0, x_0)$ and let $t^n$ be the term recursively defined as $f(f(x_n, x_n), t^{n-1})$ for $n > 0$, and $t^0 = f(a, b)$. Consider the 1-CU instance $\{F(a) \doteq s, F(b) \doteq t^n\}$, where $x_1, \ldots, x_n$ are pairwise different first-order variables and $F$ is a the context variable. This instance has an exponential number of solutions, and [7] will take exponential time. However, our algorithm solves this class of problems in polynomial time.

This paper extends the results of [7] by providing polynomial time solutions to several interesting classes of the one context unification problem such as, classes containing left- and right-ground instances, a class containing disjoint first-order variables on the two sides, and a class containing a constant number of equations. In fact, we do not have a class of examples for which our procedure takes exponential time.

Our algorithm can actually be seen as following a "divide and conquer" paradigm and it relies on polynomial time algorithms for some base cases. One of the most important base cases are instances of the form $\{F(r_1) \doteq s, F(r_2) \doteq t\}$, where the context variable $F$ does not occur in neither $r_1$ nor $r_2$. We call this the 2-restricted 1-CU problem. The algorithm for this case is presented in a companion paper [8], and while there is some overlap in the technical development with [8], the results here are disjoint from the result in [8] and they non-trivially build upon them. It was infeasible to coherently include the results in [8] in this paper.

## II. Preliminaries

We assume a fixed ranked alphabet $\mathcal{F}$ and a set of variables $\mathcal{X}$ containing first-order variables and exactly one context variable. We denote the context variable by $F$ and first-order variables by $x$ with possible subindexes. Our algorithm introduces fresh first-order variables from a set $\mathcal{Y}$, which we denote by $y$ with possible subindexes. We denote $\mathcal{X} \cup \mathcal{Y}$ as $\mathcal{V}$ and use $z$ to denote variables in $\mathcal{V}$. We will argue about terms in $\mathcal{T}(\mathcal{F}, \mathcal{X})$, $\mathcal{T}(\mathcal{F}, \mathcal{V})$ and $\mathcal{T}(\mathcal{F}, \mathcal{X}) \cup \mathcal{Y}$, and contexts in $\mathcal{C}(\mathcal{F}, \mathcal{X})$, $\mathcal{C}(\mathcal{F}, \mathcal{V})$ and $\mathcal{C}(\mathcal{F}, \mathcal{X}) \cup \mathcal{Y}$.

With $<$ we denote the prefix relation among positions and with $\prec$ the subterm relation among terms. We also define, for a term $t = \alpha(t_1, \ldots, t_n)$, $\mathtt{topsymbol}(t) = \alpha$.

In this work, we deal with *multiequations on terms*, denoted by $m$ with possible subindexes. Given a multiequation $m = (t_1 \doteq \ldots \doteq t_n)$, we call the set $\bigcup_i^n \{t_i\}$ the *top terms of* $m$, denoted $\mathtt{topterms}(m)$. Similarly, for a multiset $\Delta$ of multiequations, $\mathtt{topterms}(\Delta)$ denotes $\bigcup_{m \in \Delta}(\mathtt{topterms}(m))$. Similarly, $\mathtt{topvars}(m) = \mathtt{topterms}(m) \cap \mathcal{V}$. We also extend $\mathtt{topsymbols}$ from terms to multiequations as $\mathtt{topsymbols}(m) = \bigcup_{t \in \mathtt{topterms}(m)}(\mathtt{topsymbols}(t))$. By $|\Delta|$ we denote the number of multiequations in $\Delta$. Moreover, we consider our multiequations to be ordered and use $m[i]$ to refer to $t_i$, for every $i \in \{1, \ldots, n\}$.

A *substitution*, denoted by $\sigma, \theta, \eta$, is a total function $\sigma : \mathcal{V} \to \mathcal{T}(\mathcal{F}, \mathcal{V}) \cup \mathcal{C}(\mathcal{F}, \mathcal{V})$ such that $\alpha\sigma \in \mathcal{T}(\mathcal{F}, \mathcal{V})$ if $\alpha$ is a first-order variable and $\alpha\sigma \in \mathcal{C}(\mathcal{F}, \mathcal{V})$ if $\alpha$ is a context variable.

The *domain of a substitution* $\sigma$, denoted $\mathtt{dom}(\sigma)$, is defined as usual, i.e. $\mathtt{dom}(\sigma) = \{z \in \mathcal{V} \mid z\sigma \neq z\}$. The *composition* of $\sigma$ and $\theta$, denoted $\theta \circ \sigma$, is defined as $\{\alpha \mapsto \alpha\sigma\theta \mid \alpha \in \mathtt{dom}(\sigma) \cup \mathtt{dom}(\theta)\}$.

For substitutions $\sigma, \theta$, $\sigma = \theta$ holds if $\forall z \in \mathcal{Y} : z\sigma = z\theta$. Moreover, $\sigma$ is *more general* than $\theta$, denoted $\sigma \leq \theta$, if there exists $\eta$ such that $\sigma = \theta \circ \eta$.

Substitutions are extended to be mappings from terms to terms, i.e. $\sigma : \mathcal{T}(\mathcal{F}, \mathcal{V}) \to \mathcal{T}(\mathcal{F}, \mathcal{V})$, as $g(t_1, \ldots, t_n)\sigma = g(t_1\sigma, \ldots, t_n\sigma)$ and $F(t)\sigma = F\sigma(t\sigma)$. In addition, substitutions are also extended, in a similar way, to be mappings from contexts to contexts, i.e. $\sigma : \mathcal{C}(\mathcal{F}, \mathcal{V}) \to \mathcal{C}(\mathcal{F}, \mathcal{V})$. We also define $m\sigma = (t_1\sigma \doteq \ldots \doteq t_n\sigma)$, for a multiequation $m = (t_1 \doteq \ldots \doteq t_n)$, $\Delta\sigma = \biguplus_{m \in \Delta}(m\sigma)$, for a multiset of multiequations $\Delta$, and $L\sigma = \langle t_1\sigma, \ldots, t_n\sigma \rangle$, for a list of terms $L = \langle t_1, \ldots, t_n \rangle$.

A *unifier* of two terms $s, t$ is a substitution $\sigma$ such that $s\sigma = t\sigma$. A unifier does not always exist. We capture that situation by simply saying that the unifier of $s$ and $t$ is $\bot$. We define the *most general unifier* of terms $s$ and $t$, denoted $\mathtt{mgu}(s = t)$, as *any* substitution $\sigma$ such that, for every unifier $\theta$ of s and t, $\sigma \leq \theta$ holds. If such substitution does not exist we say that $\mathtt{mgu}(s = t)$ is *not defined*, denoted $\mathtt{mgu}(s = t) = \bot$. In an abuse of notation, we assume that $t\sigma = \bot$ for every term $t$ if $\sigma = \bot$ and extend the definitions for the application of a substitution on a term, multiequation, multiset of multiequations, and list of terms accordingly. Moreover, $\mathtt{mgu}$ is extended to multiequations in the natural way.

Although 1-CU is defined as a set of equations over terms in $\mathcal{T}(\mathcal{F}, \mathcal{X})$, we can always transform an 1-CU instance (by several applications of the usual decomposition operation) to the following more restricted definition without loss of generality.

**Definition II.1** (1-CU)**.** *An instance $\mathcal{I}$ of the 1-CU problem is a set of equations $\{F(s_1) \doteq t_1, \ldots, F(s_n) \doteq t_n\}$, where $\forall i \in \{1, \ldots, n\} : \mathtt{topsymbol}(t_i) \neq F$. A solution of $\mathcal{I}$ is a substitution $\sigma$ such that $F(s_i)\sigma = t_i\sigma$, for every $i \in \{1, \ldots, n\}$.*

**Definition II.2** (size)**.** *Let $\mathcal{I}$ be a 1-CU instance. The size of $\mathcal{I}$, denoted $||\mathcal{I}||$, is defined as the sum of the sizes of the terms in the equations of $\mathcal{I}$.*

We assume the DAG representation for terms and hence the size $||\mathcal{I}||$ is just the number of nodes in the DAG representing all terms in $\mathcal{I}$. In the sequel, we assume this measure for 1-CU instances and hence we do not state it explicitly every time we refer to a *polynomial time* algorithm. Moreover, when we state that a unification problem *can be solved* in polynomial time, we refer to both its decisional version (deciding unifiability) and functional version (finding a unifier).

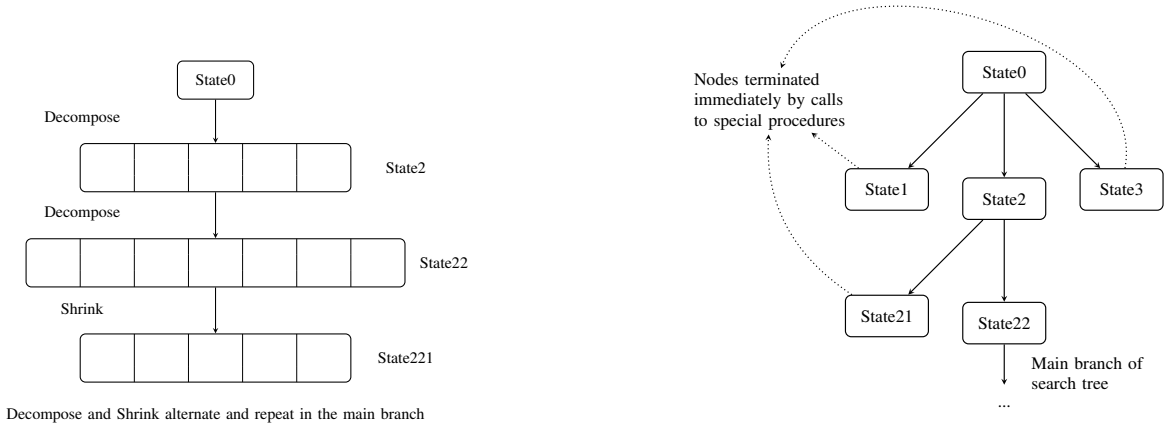


Fig. 1. Visualizing the Procedure

## III. Overview of the Procedure

In this section, we present a very high-level overview of our procedure for solving the one context unification problem.

Our procedure builds a search tree starting from an initial state that encodes the input problem. We present inference rules that can be used to generate child states from a parent state. As shown in the illustration in Figure 1 (left), a state can be decomposed to give a new state that encodes several subproblems that together solve the original problem. Repeated decomposition steps can yield larger and larger number of subproblems. To keep the number of subproblems polynomially bounded, we have *shrink* rules that eliminate some subproblems and reduce the total number of subproblems contained in one state (see Figure 1).

Some inference rules can potentially create multiple children (as shown in Figure 1 right). However, in all such cases, there will be exactly one child on which the inference rules are recursively applied. All other children are solved by specialized (polynomial-time) procedures immediately. This keeps the complexity of our procedure in polynomial time.

There are two possible outcomes when calling a specialized procedure. Either the specialized procedure is successful in finding unifiers or it concludes there is no unifier (for the subproblem). If the specialized procedure is successful, then the whole search terminates. If not, the search continues along the main branch.

## IV. Special Cases

In this section we present particular cases of 1-CU that can be solved in polynomial time. Our procedure relies on these particular cases. In particular, these particular cases will help in either shrinking the main branch of our search tree, or immediately terminate the side branches of our search tree.

**Claim IV.1** (Clash)**.** *Let $\mathcal{I}$ be a 1-CU instance of the form $\mathcal{I} = \mathcal{I}' \cup \{F(u_1) \doteq f(s_1, \ldots, s_m), F(u_2) \doteq g(t_1, \ldots, t_{m'})\}$, with $f \neq g$. Then, $\mathcal{I}$ can be solved in polynomial time and every solution $\sigma$ satisfies $F\sigma = [\bullet]$.*

*Proof.* The fact that every solution $\sigma$ must satisfy $F\sigma = [\bullet]$ directly follows from $f \neq g$. Hence, this particular case reduces to the first-order unification problem $\mathcal{I}' = \mathcal{I}\{F \to \bullet\}$, which can be solved in polynomial time w.r.t. $||\mathcal{I}||$. □

Every unifier for a 1-CU problem maps $F$ to some context $C[\bullet]$. The position of the hole $\bullet$ in $C[\bullet]$ is called the *hole position* of $C$ and denoted by $\mathtt{hp}(C)$. A unifier $\sigma$ of a 1-CU problem can be found immediately once we fix $\mathtt{hp}(F\sigma)$: once $\mathtt{hp}(F\sigma)$ is fixed, a 1-CU problem reduces to a first-order unification problem.

Next, consider another special case where we have one equation that contains $F$ on both sides. This special case was solved in a previous paper [7]. In the case of the equation $F(s) \doteq C[F(t)]$, one of the key observations is that the hole position of every context that is a solution for $F$ has to be a prefix or exponentiation of $\mathtt{hp}(C)$.

**Theorem IV.1** ([7])**.** *Let $\mathcal{I}$ be a 1-CU instance of the form $\mathcal{I} = \mathcal{I}' \cup \{F(s) \doteq C[F(t)]\}$, where $C$ is a non-empty context. Then, $\mathcal{I}$ can be solved in polynomial time.*

Finally, consider the case when we have equations $F(u_1) \doteq s$ and $F(u_2) \doteq C[s]$ and $C$ is nonempty.

**Claim IV.2** (cyclic)**.** *Let $\mathcal{I}$ be a 1-CU instance of the form $\mathcal{I} = \mathcal{I}' \cup \{F(u_1) \doteq s, F(u_2) \doteq C[s]\}$, where $C$ is a non-empty context. Then, $\mathcal{I}$ can be solved in polynomial time.*

*Proof.* Consider the instance $\mathcal{I}'' = \mathcal{I} \cup \{F(u_2) \doteq C[F(u_1)]\}$. The lemma follows from the fact that $\mathcal{I}$ and $\mathcal{I}''$ have the same set of solutions, $||\mathcal{I}''||$ is polynomial w.r.t. $||\mathcal{I}||$, and $\mathcal{I}''$ can be solved in polynomial time by Theorem IV.1. □

## V. Inference Rules for One Context Unification

We present inference rules for solving the one context unification problem in this section.

### A. Defining the State

Our inference rules operate on states (configurations), which are pairs of the form

$$\langle \Delta, L \rangle$$

where $\Delta$ is a multiset of multiequations and $L$ is a list of terms. Given a 1-CU instance $\{F(s_1) \doteq t_1, \ldots, F(s_n) \doteq t_n\}$, the initial state of our algorithm is

$$\langle \{\!\{ t_1 \doteq \ldots \doteq t_n \}\!\}, \langle s_1, \ldots, s_n \rangle \rangle$$

Our states preserve the invariant that each multiequation in $\Delta$ has $|L|$ top terms and the value $|L|$ remains unchanged. Hence, if we start with the above initial state, then for every state $\langle \Delta, L \rangle$ generated by our inference rules, we will have that
(a) $|L| = n$ and
(b) the multiequations in $\Delta$ have $n$ top terms.
In the sequel, especially in the section proving correctness of our procedure, whenever we say *state*, we implicitly assume that it satisfies (a) and (b).

Since $\Delta$ is a multiset, we fix the following convention: whenever we refer to a multiequation $m \in \Delta$, we will mean *one specific* element of $\Delta$ and not all the multiple occurrences of the element $m$ in $\Delta$. We will later see that our inference rules will guarantee that $\Delta$ always turns into a set, and it is a multiset only in "transient" states.

### B. Mapping State to 1-CU Instances

We will next formally define a mapping from states to 1-CU instances. This mapping will help in stating the soundness and completeness of the inference rules.

**Definition V.1.** *Given a multiset of multiequations $\Delta$ and $m \in \Delta$, we define the* context unifier of $m$ in $\Delta$, *denoted* $\mathtt{cmgu}(m, \Delta)$ *(or simply* $\mathtt{cmgu}(m)$ *when $\Delta$ is clear from the context), as* $\mathtt{mgu}(\Delta \setminus \{\!\{m\}\!\})$.

*Similarly, let $\theta = \mathtt{mgu}(m)$. We define $\Delta|_{\widehat{m}}$ as $(\Delta \setminus \{\!\{m\}\!\})\theta$.*

An intermediate state $\langle \Delta, L \rangle$ of our procedure spans $|\Delta|$ 1-CU instances.

**Definition V.2.** *Let $S = \langle \Delta, \langle u_1, \ldots, u_n \rangle \rangle$ be a state, let $m \in \Delta$ be a multiequation, and let $\theta = \mathtt{cmgu}(m)$. We define the 1-CU instance spanned by $m$ in $S$, denoted $P(m, S)$ (or simply $P(m)$ if $S$ is clear from the context), as*

$$P(m) = \begin{cases} \bigcup_{i \in \{1,\ldots,n\}} \{F(u_i\theta) \doteq (m[i])\theta\} & \text{if } \theta \neq \bot \\ \bot & \text{otherwise.} \end{cases}$$

*where $m[i]$ denotes the $i$-th top term in the multiequation $m$.*

Note that several copies of the same multiequation in $\Delta$ span the same problem. Our Shrinking rules appropriately handle such situation to avoid solving the same subproblem twice.

### C. The ForcedDecompose Rule

A useful observation to understand our procedure is that, given a solution $\sigma$ to a 1-CU instance $\{F(s_1) \doteq t_1, \ldots, F(s_n) \doteq t_n\}$, $\mathtt{hp}(F\sigma)$ completely characterizes $\sigma$. Roughly speaking, to recover $\sigma$ from $\mathtt{hp}(F\sigma)$, it is enough to (i) unify all terms in $t_1, \ldots, t_n$ at positions disjoint with $\mathtt{hp}(F\sigma)$ and (ii) solve a first-order unification problem. Hence, a procedure for 1-CU may proceed by guessing $\mathtt{hp}(F\sigma)$. Our procedure builds on that idea by incrementally constructing $\mathtt{hp}(F\sigma)$. However, the cases where a guess is unavoidable are handled carefully to avoid considering exponentially many cases.

Recall the decomposition rule for first-order unification which replaces the equation $f(s,t) \doteq f(u,v)$ by two equations $s \doteq u$ and $t \doteq v$. Our algorithm uses a variant of this usual term decomposition inference rule. This rule is applied on the multiequations in $\Delta$. Such decomposition leads to a multiset of multiequations. For example, consider the following 1-CU instance:

$$\{F(s_0) \doteq fxx,\ F(s_1) \doteq fu_1v_1,\ \ldots,\ F(s_n) \doteq fu_nv_n\}$$

where we have removed braces and written $f(u_1, v_1)$ as $fu_1v_1$.

As mentioned above, this 1-CU instance corresponds to the following initial state of our algorithm:

$$S_0 = \langle \{\!\{ fxx \doteq fu_1v_1 \doteq \ldots \doteq fu_nv_n \}\!\}, \langle s_0, \ldots, s_n \rangle \rangle$$

Checking whether $\mathtt{hp}(F\sigma) = \lambda$ (or, equivalently $F\sigma = [\bullet]$) is a solution reduces to first-order unification. If we can find a solution $\sigma$ where $F\sigma = [\bullet]$, we are done. If not, then we should consider the cases where $|\mathtt{hp}(F\sigma)| > 0$. After decomposition, we obtain the following state:

$$S_1 = \langle \{\!\{ x \doteq u_1 \doteq \ldots \doteq u_n,\ x \doteq v_1 \doteq \ldots \doteq v_n, \}\!\}, \langle s_0, \ldots, s_n \rangle \rangle$$

This state, according to Definition V.2, spans the following two 1-CU instances:

1) $\{F(s_0\sigma_1) \doteq x\sigma_1, F(s_1\sigma_1) \doteq u_1\sigma_1, \ldots, F(s_n\sigma_1) \doteq u_n\sigma_1, \}$, where $\sigma_1 = \mathtt{mgu}(x \doteq v_1 \doteq \ldots \doteq v_n)$, and
2) $\{F(s_0\sigma_2) \doteq x\sigma_2, F(s_1\sigma_2) \doteq v_1\sigma_2, \ldots, F(s_n\sigma_2) \doteq v_n\sigma_2, \}$, where $\sigma_2 = \mathtt{mgu}(x \doteq u_1 \doteq \ldots \doteq u_n)$.

Note that (1) corresponds to the cases where $1 \leq \mathtt{hp}(F\sigma)$, and (2) corresponds to the case where $2 \leq \mathtt{hp}(F\sigma)$ and hence our state does not miss any solution of the original problem where $|\mathtt{hp}(F\sigma)| > 0$. However, considering both cases above and solving the corresponding subproblem *separately* is not a good idea, since the algorithm may end up exploring too many equivalent possibilities. In fact, this is one reason for the exponential running time of the algorithm presented in [7], which is relevant the class of instances presented in the introduction.

A possible alternative is to "delay" the computation and application of the substitutions $\sigma_1$ and $\sigma_2$ as much as possible. But, we can not apply the usual *decomposition* rule on state $S_1$ since both multiequations in it have a variable $x$. So, how do we make progress? Here we use our *first key idea*: we extend

the decomposition rule to allow minimal instantiation of $x$ that will allow us to progress (with decomposition steps).

Concretely, let us assume that the $u_i$'s and $v_i$'s of our example are of the forms $u_i = f(u_i^1, u_i^2)$ and $v_i = f(v_i^1, v_i^2)$. In this case, note that $\sigma_1$ would have instantiated $x$ by a term of the form $f(\_,\_)$ and $\sigma_2$ would also have instantiated $x$ by a (possibly different) term of the *same form*. Hence, we can keep both options open for $x$ in the next state by instantiating $x$ in terms of *fresh* variables $y_1, y_2$. That is, we apply a substitution of the form $x \to f(y_1, y_2)$, where $f$ is uniquely determined by $\Delta$. This allows us to proceed with decompose and avoiding committing to any one branch. We call this inference rule the `ForcedDecompose` rule, and it is shown in Figure 3. Unfortunately, the `ForcedDecompose` rule adds new variables to our problem.

After applying the `ForcedDecompose` rule, we get the following state in our example:

$$\begin{gathered} S_2 = \langle \{\!\{ y_1 \doteq u_1^1 \doteq \ldots \doteq u_1^1, \\ y_2 \doteq u_1^2 \doteq \ldots \doteq u_n^2, \\ y_1 \doteq v_1^1 \doteq \ldots \doteq v_n^1, \\ y_2 \doteq v_1^2 \doteq \ldots \doteq v_n^2 \}\!\}, \\ \langle s_0\{x \to f(y_1, y_2)\}, \ldots, s_n\{x \to f(y_1, y_2)\}\rangle\rangle \end{gathered}$$

Our approach consists of applying this lazy instantiation followed by term decomposition (the `ForcedDecompose` rule), but ensuring that, every time we apply this rule, we decrease the measure $|\mathtt{subterms}(\mathtt{topterms}(\Delta)) \setminus \mathcal{Y}|$, i.e. the total number of terms occurring in the multiequations of $\Delta$ that are not fresh variables. If we apply the `ForcedDecompose` rule arbitrarily, then the above measure may not decrease. Even in regular first-order unification, a decomposition step is not guaranteed to remove some subterm from $\Delta$.

Here we use our *second key idea*: instead of decomposing arbitrarily any/all multiequations in $\Delta$, we enforce that the above measure decreases by decomposing only a non-empty submultiset $\Gamma \subseteq \Delta$ of multiequations at a time. The selected submultiset $\Gamma$ satisfies that every term in $\mathtt{topterms}(\Gamma)$ is maximal with respect to term inclusion in $\Delta$. Note that the same idea can be used in usual first-order unification: by decomposing on maximal terms, one can guarantee that those terms will get removed from $\Delta$. A consequence of this side condition in our `ForcedDecompose` rule is that the variables in $\mathcal{Y}$ will never occur in the multiequations in $\Delta$ as subterms of other terms, i.e., we will maintain the invariant that $\mathtt{topterms}(\Delta) \subset (\mathcal{T}(\Sigma, \mathcal{X}) \cup \mathcal{Y})$.

We formally define the `ForcedDecompose` rule next. First, let us recall a variant of the traditional term decomposition rule for multiequations.

**Definition V.3.** *Let $\Delta$ be a set of multiequations and let $m = f(s_1^1, \ldots, s_l^1) \doteq \ldots \doteq f(s_1^k, \ldots, s_l^k)$ be a multiequation in $\Delta$ such that $\mathtt{topvars}(m) = \emptyset$ and $\mathtt{topsymbols}(m) = \{f\}$, with $l = \mathtt{ar}(f)$.*

- *By $\mathtt{Decompose}(m)$ we denote the multiset of multiequations $\bigcup_{i=1}^{l}(\{s_i^1 \doteq \ldots \doteq s_i^k\})$,*
- *by $\mathtt{Decompose}(\Delta, m)$ we denote $\Delta \setminus \{m\} \cup \mathtt{Decompose}(m)$, and*
- *`Decompose` is also extended to a submultiset of multiequations $\Gamma \subseteq \Delta$ as $\mathtt{Decompose}(\Delta, \Gamma) = (\Delta \setminus \Gamma) \cup \bigcup_{m \in \Gamma} \mathtt{Decompose}(m)$.*

**Definition V.4.** *Let $S = \langle \Delta, L\rangle$ be a state of the algorithm, let $\Gamma \subseteq \Delta$ be a subset of multiequations such that $\mathtt{topsymbols}(\Gamma) = \{f\}$, and let $\mathcal{Y}$ be a set of first-order variables disjoint with $\mathtt{vars}(S)$. Let $\{x_1, \ldots, x_k\}$ be $\mathtt{topvars}(\Gamma)$ and let $\sigma$ be $\bigcup_{i=1}^{k}(\{x_i \to f(y_1^i, \ldots, y_{\mathtt{ar}(f)}^i)\})$, where $y_j^i \in \mathcal{Y}$, for $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \mathtt{ar}(f)\}$.*

*Then, $\mathtt{ForcedDecompose}(S, \Gamma, \mathcal{Y})$ is defined as the state $\langle (\Delta \setminus \Gamma) \cup \mathtt{Decompose}(\Delta\sigma, \Gamma\sigma), L\sigma\rangle$.*

Note that the substitution $\sigma$ is not applied to $\Delta \setminus \Gamma$ in the definition of $\mathtt{ForcedDecompose}(S, \Gamma, \mathcal{Y})$. This will not be a problem, due to the conditions for the application of rule `ForcedDecompose`; see Figure 3.

### *D. The Shrinking Rules*

If we only rely on decomposition, we will end up with exponentially many multiequations in $\Delta$. To avoid this explosion, we exhaustively apply a sequence of *shrinking operations* to $\Delta$ before applying every decomposition step. Such shrinking rules are shown in Figure 2. The shrinking rules simplify the current state $\langle \Delta, L\rangle$ of the algorithm by either completely solving one of the problems spanned by $\langle \Delta, L\rangle$ in polynomial time (rules `CycleOrClash` and `TwoNonVar`), or applying substitutions that preserve all solutions (rules `InvEq`, `NoSol`).

A crucial property of our algorithm (captured in Lemma V.18) is that, if none of the shrinking rules can be applied, then $|\Delta|$ is (bounded by) a polynomial function of $|\mathtt{subterms}(\mathtt{topterms}(\Delta)) \setminus \mathcal{Y}|$. This fact, together with the fact that every application of `ForcedDecompose` reduces $|\mathtt{subterms}(\mathtt{topterms}(\Delta)) \setminus \mathcal{Y}|$ and the application of the other rules does not increase that value, completes our termination argument, by induction on $|\mathtt{subterms}(\mathtt{topterms}(\Delta)) \setminus \mathcal{Y}|$.

We will now describe the shrinking inference rules. The first three rules each remove a multiequation $m$ from the multiset $\Delta$. The last rule will simplify the problem by applying a substitution. If $P(m)$ (Definition V.2) has no solution, then we can delete $m$ from $\Delta$. In what cases can we easily deduce that $P(m)$ has no solution?

*a) The `NoSol` inference rule:* If $\Delta - \{\!\{m\}\!\}$ is *not* unifiable, then it means that hole position of $F$ can not lie at (or below) position corresponding to $m$. Hence, we can remove $m$. To remove $m$, we have to unify $m$, apply the unifier to all other multiequations in $\Delta$ and continue. This is captured in the `NoSol` rule in Figure 2, which states that if there is a multiequation $m \in \Delta$ such that its context unifier (Definition V.1) is $\bot$, then we can remove $m$ from $\Delta$ and update $\Delta$ to $\Delta|_{\widehat{m}}$. Recall that if $\theta = \mathtt{mgu}(m)$, then $\Delta|_{\widehat{m}}$ is defined as $(\Delta \setminus \{\!\{m\}\!\})\theta$.

*b) The `CycleOrClash` inference rule:* If the 1-CU instance $P(m)$ has two equations of the form $F(u_1) = f(\ldots)$ and $F(u_2) = g(\ldots)$ for some $f \neq g$, then we can easily determine if $P(m)$ has a solution using Claim IV.1. Similarly,

if the 1-CU instance $P(m)$ has two equations of the form $F(u_1) = s$ and $F(u_2) = C[s]$ for some nonempty context $C$, then we can easily determine if $P(m)$ has a solution using Claim IV.2. In both these cases, if $P(m)$ has a solution, we can terminate the search and report success. If $P(m)$ has no solution, then we can remove $m$ from $\Delta$. This process is formalized in the `CycleOrClash` inference rule.

*c) The* `TwoNonVar` *inference rule:* Let us assume that we have access to an *oracle* that can solve problems $P(m)$ that have at most two different non-variable terms on the right-hand side. The inference rule `TwoNonVar` uses this oracle to remove such branches. This rule works in the same way as the rule `CycleOrClash`.

*d) The* `InvEq` *inference rule:* Rather than removing multiequations $m$ from $\Delta$, our last shrinking rule `InvEq` removes variables from the state by applying substitutions that can be deduced to hold in every branch of the search tree. How to find such "globally" valid substitutions? We need a few definitions for this purpose.

**Definition V.5.** *Let $\Delta$ be a multiset of term multiequations. Let $\Delta^\ell$ be the set obtained by marking each occurrence of a multiequation $m$ in $\Delta$ with a different mark*[1]*. The graph $G(\Delta)$ is defined as the undirected graph that has $\texttt{topterms}(\Delta) \cup \Delta^l$ as the nodes and the relation $\{(s, m^l) \mid m^l \in \Delta^l, s \in \texttt{topterms}(m)\}$ as the edges.*

Cycles in the graph $G(\Delta)$ are special: if terms $s, t$ lie on a cycle, then every solution $\sigma$ of every problem $P(m)$ should unify $s$ and $t$.

**Definition V.6.** *Let $\Delta$ be a set of multiequations and let $s, t \in$ `topterms` be distinct terms. We say that $\Delta$* induces the equality *$s = t$, denoted $\Delta \models (s = t)$, if $s$ and $t$ lie on some cycle in $G(\Delta)$.*

It follows directly from the definition above that equations induced by $\Delta$ can be computed efficiently.

**Lemma V.7.** *Given a multiset $\Delta$ of multiequations, terms $s, t \in$ `topterms` such that $\Delta \models (s = t)$, if they exist, can be found in polynomial time with respect to $|\Delta|$.*

If $s = t$ is an induced equality, then every solution $\sigma$ of the 1-CU problem should make $s\sigma = t\sigma$, and hence, we can unify $s$ and $t$ and apply the unifier to our state without losing any solutions. This action is performed by the `InvEq` rule.

Note that all the shrinking rules rely on the application of a most general unifier. Such unifier $\theta$ does not necessarily always exist and hence, by our convention mentioned in Section II, in that case $\theta = \bot$ and $\langle \Delta\theta, L\theta \rangle = \langle \bot, \bot \rangle$. This allows to simplify the presentation by having a single failing rule `Fail` (Figure 3) while still making the failing cases explicit.

**Remark V.8.** *If $\Delta$ has two copies of the same multiequation $m$, say $s_1 \doteq \cdots \doteq s_n$, then we have $\Delta \models (s_1 = \cdots = s_n)$. The* `InvEq` *can be used to unify all the $s_i$'s in the state, and as a result the problem $P(m)$ has all right-hand side terms equal. The inference rule* `TwoNonVar` *can remove such $m$. Hence, if the multiset $\Delta$ has multiple occurrences of $m$, it can get rid of the copies and get converted to a set.*

[1]If $\Delta$ has $m$ twice in it, then the set $\Delta^\ell$ has two elements, say $m^{l_1}, m^{l_2}$.

### *E. The algorithm*

All our inference rules are presented in Figures 2 and 3. To obtain a polynomial time procedure, we will apply our inference rules according to a particular strategy. Specifically, our strategy gives priority to the shrinking rules over the `ForcedDecompose` rule and thus can be described as sequences of rule application of the form: $(\texttt{Shrink}^! \cdot \texttt{ForcedDecompose})^!$ where `Shrink` refers to (`CycleOrClash` | `TwoNonVar` | `InvEq` | `NoSol`).

Our procedure currently uses an oracle to solve some subproblems. Assuming that the oracle runs in polynomial time, to prove polynomial running time of our procedure, we have to argue that

(a) all intermediate subproblems have polynomial size,
(b) every application of `ForcedDecompose` yields polynomially many subproblems, and
(c) the derivations in our algorithm have polynomial length.

Note that, roughly speaking, to prove (a) and (b) it suffices to argue that every state $\langle \Delta, L \rangle$ considered by our algorithm can be represented in polynomial space. Regarding correctness, we must argue that our rules neither miss solutions (completeness) nor introduce new solutions (soundness).

### *F. Correctness*

Let us first extend our definition of a solution of an 1-CU instance to define solution of a state.

**Definition V.9.** *Let $\langle \Delta, L \rangle$ be a state of our procedure. A solution for $\langle \Delta, L \rangle$ is a pair $\langle m, \theta \rangle$, where $m \in \Delta$ is a multiequation such that $P(m) \neq \bot$, and $\theta$ is a solution for $P(m)$.*

**Lemma V.10.** *Let $S = \langle \{m^\ell\} \cup \Delta, L \rangle$ be a state of our algorithm. $S$ has a solution if and only if either $P(m)$ has a solution or $\langle \Delta|_{\widehat{m^\ell}}, L\sigma \rangle$ has a solution, where $\sigma = mgu(m)$.*

*Proof.* This Lemma easily follows from Definition V.2 by induction on $|\Delta|$. Note that, for any solution $\langle m', \theta \rangle$ with $m' \neq m$, $\texttt{mgu}(m) \leq \theta$. Also note that if $mgu(m) = \bot$, then $\langle \Delta|_{\widehat{m}}, L\sigma \rangle = \langle \bot, \bot \rangle$ and $\langle \Delta|_{\widehat{m}}, L\sigma \rangle$ has no solution. □

We denote an application of an inference rule using infix $\rightarrow$ notation, possibly labeled with the name of the inference rule. Correctness of rules `CycleOrClash`, `TwoNonVar`, and `NoSol`, stated in the next two lemmas, is a direct consequence of the previous lemma.

**Lemma V.11.** *Let $\langle \Delta, L \rangle \rightarrow_r (solve(P_m) \mid \langle \Delta', L' \rangle\rangle)$ be an inference step with $r \in \{\texttt{CycleOrClash}, \texttt{TwoNonVar}\}$. Then, $\langle \Delta, L \rangle$ has a solution if and only if either $P(m, \Delta)$ has a solution or $\langle \Delta', L' \rangle$ has a solution.*

**Lemma V.12.** *Let $\langle \Delta, L \rangle \rightarrow_{\texttt{NoSol}} \langle \Delta', L' \rangle\rangle$ be an inference step. Then, $\langle \Delta, L \rangle$ has a solution if and only if $\langle \Delta', L' \rangle$ has a solution.*

$$\texttt{TwoNonVar:}\quad \frac{\langle \Delta, L\rangle}{Solve(P(m)) \mid \langle \Delta|_{\widehat{m}}, L\texttt{mgu}(m)\rangle} \quad \exists m \in \Delta : |\texttt{topnonvars}(m\texttt{cmgu}(m))| \leq 2$$

$$\texttt{CycleOrClash:}\quad \frac{\langle \Delta, L\rangle}{Solve(P(m)) \mid \langle \Delta|_{\widehat{m}}, L\texttt{mgu}(m)\rangle} \quad \exists m \in \Delta : P(m) \text{ satisfies the conditions of Lemma IV.2 or Lemma IV.1.}$$

$$\texttt{InvEq:}\quad \frac{\langle \Delta, L\rangle}{\langle \Delta\sigma, L\sigma\rangle} \quad \Delta \models (s = t) \text{ and } \sigma = \texttt{mgu}(s \doteq t)$$

$$\texttt{NoSol:}\quad \frac{\langle \Delta, L\rangle}{\langle \Delta|_{\widehat{m}}, L\texttt{mgu}(m)\rangle} \quad \exists m \in \Delta : \texttt{cmgu}(m) = \bot$$

Fig. 2. Shrinking rules of the 1-CU algorithm

$$\texttt{ForcedDecompose:}\quad \frac{\langle \Delta = \Gamma \cup \Delta', L\rangle}{solve(P(m_1)\sigma) \mid \ldots \mid solve(P(m_{|\Delta|})\sigma) \mid \texttt{ForcedDecompose}(\langle \Delta, L\rangle, \Gamma, \mathcal{Y})}$$

$|\texttt{topsymbols}(\Gamma)| = 1$ and
every $t \in \texttt{topterms}(\Gamma)$ is maximal in $\Delta$
w.r.t. term inclusion and $\sigma = \{F \to [\bullet]\}$.

$$\texttt{Fail:}\quad \frac{\langle \bot, \bot\rangle}{fail}$$

Fig. 3. Decomposition and failing rules of the 1-CU algorithm

*Proof.* Note that $P(m) = \bot$, where $m \in \Delta$ is the multiequation that satisfies the condition for the application of NoSol. □

**Lemma V.13.** *Let* $\langle \Delta, L\rangle \to_{\texttt{ForcedDecompose}} (solve(P(m_1)\sigma) \mid \ldots \mid solve(P(m_{|\Delta|})\sigma) \mid \langle \Delta', L'\rangle)$ *be an inference step. Then,* $\langle \Delta, L\rangle$ *has a solution* $\langle m, \theta\rangle$ *if and only if either* $F\theta = [\bullet]$ *and the first-order unification problem* $P(m)\sigma$ *has a solution or* $\langle \Delta', L'\rangle$ *has a solution.*

*Proof.* Note that, if $\langle \Delta, L\rangle$ has a solution $\langle m, \theta\rangle$ then either $\theta$ maps $F$ to $[\bullet]$ or it does not. The former case is checked by solving the first-order unification problem $P(m)\sigma = P(m)\{F \mapsto [\bullet]\}$, whereas the latter case implies that $\langle \Delta', L'\rangle$ has a solution, since the ForcedDecompose operation simply decomposes some multiequations without making any assumptions. For the other implication, note that the ForcedDecompose operation is only defined if the top symbol in all the non-variable terms in the decomposed multiequations are the same, and hence this rule application does not introduce new solutions. □

**Lemma V.14.** *Let* $\langle \Delta, L\rangle \to_{\texttt{InvEq}} \langle \Delta', L'\rangle\rangle$ *be a derivation in our algorithm. Then,* $\langle \Delta, L\rangle$ *has a solution if and only if* $\langle \Delta', L'\rangle$ *has a solution.*

*Proof.* The correctness of rule InvEq follows from Definition V.6 and the fact that $\Delta \models (s = t)$ holds for some state $S = \langle \Delta, L\rangle$ if and only if $\texttt{mgu}(s, t) \leq \theta$, for any solution $\langle m, \theta\rangle$ of $S$. □

We conclude that each inference rule preserves unifiability. We next show *progress*; that is, any derivation starting from an initial state either terminates early (with success) or it reaches a "terminal" state, if we apply the rules exhaustively. Proof of the following lemma can be found in the appendix.

**Lemma V.15.** *Let* $S = \langle \Delta, L\rangle$ *be a state of our procedure such that no rule can be applied to* $S$*. Then,* $\Delta = \emptyset$*.*

The correctness of the algorithm follows from the previous lemmas.

**Theorem V.16.** *The algorithm is correct regardless of the rule application strategy.*

Now we need to establish the procedures complexity. As seen in the example of Section I-A, a rule application strategy is required to guarantee termination.

### G. Runtime analysis

Henceforth we assume that our algorithm applies the rules of Figures 2 and 3 according to the strategy $(\texttt{Shrink}^! \cdot \texttt{ForcedDecompose})^!$, where Shrink is the collection of all four shrinking rules. Let us remark that we assume that terms are represented using a directed acyclic graphs (DAG). The size of a DAG is defined as its number of nodes. We assume that all the terms (including subterms) involved in our problem


Fig. 4. An example acyclic graph with an unbounded number of $m$-nodes (red nodes), where each red node has two neighbours that can both reach some $f$-node (blue node) (using light blue $v$-nodes).

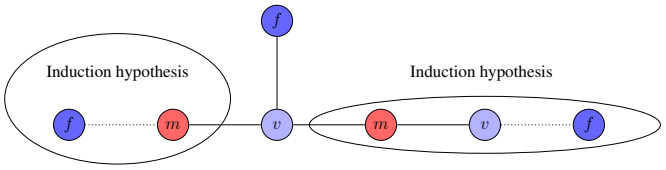


Fig. 5. Proof of induction step: the node $m$ in the middle is removed from the graph to obtain at least two disjoint graphs shown in the left and right.

are represented as nodes in a single DAG $D$. Without loss of generality, we assume that $D$ is minimal in size, and hence the correspondence between terms and nodes is bijective. Hence, We can then refer to nodes of $D$ and subterms of the problem as if they were the same thing. A crucial observation is that the *number* of terms represented in DAG is preserved by the application of substitutions resulting from the unification of terms of the DAG. This is because application of such substitutions can be achieved by manipulating only the *edges* of the DAG, leaving its nodes untouched.

The only source of difficulty is that our procedure introduces fresh variables from a set $\mathcal{Y}$, and hence the DAG $D$ will grow. However, in this section we prove a bound on the size of $D$ that does not depend on the number of introduced variables. In the rest of the paper we will not refer to $D$, We instead prove that our algorithm is polynomial w.r.t. the total number of different subterms occurring in the input $\mathcal{I}$, which is the same as the size of the DAG representing $\mathcal{I}$.

The following property, mentioned in the previous sections, is related to the previous observation about the DAG representation for terms: The algorithm does not increase the total number of subterms in the multiequations of $\Delta$ that are not freshly introduced variables from $\mathcal{Y}$. Moreover, the `ForcedDecompose` rule strictly reduces this measure. The maximality of $\Gamma$ in the conditions for `ForcedDecompose` is crucial here.

**Lemma V.17.** *Let* $\langle \Delta_0, L_0 \rangle \to^* \langle \Delta_k, L_k \rangle$ *be a derivation starting from a valid initial state. Then,* $|\mathtt{subterms}(\mathtt{topterms}(\Delta_k) \setminus \mathcal{Y})| \leq |\mathtt{subterms}(\mathtt{topterms}(\Delta_0) \setminus \mathcal{Y})|$. *Moreover, if* $\langle \Delta_0, L_0 \rangle \to^* \langle \Delta_k, L_k \rangle \to_{\mathtt{ForcedDecompose}} \langle \Delta_{k+1}, L_{k+1} \rangle\rangle$ *is a derivation from a valid initial state, then,* $|\mathtt{subterms}(\mathtt{topterms}(\Delta_{k+1})) \setminus \mathcal{V}| < |\mathtt{subterms}(\mathtt{topterms}(\Delta_k)) \setminus \mathcal{V}|$.

One of the key facts about our inference system is that the cardinality of $\Delta$ in any state generated in a derivation is polynomially bounded. The main part of the proof can be explained as a puzzle: given $n$ blue nodes, assume we have to construct a bipartite graph by adding any number of red nodes and any number of light blue nodes with the following constraints: (a) the graph is acyclic, (b) all red nodes are in one partition and the blue and light blue nodes are in the other partition, and (c) each red node has 3 neighbours, and there is a path from each neighbour to a blue node. The problem is to find a bound on the maximum possible number of red nodes that one can add. As part of the proof of the lemma below, we prove a quadratic bound for the above puzzle. Note that if each red node is required to have only 2 neighbours (with the same property), then the number of red nodes can be unbounded, as demonstrated in the graph in Figure 4.

**Lemma V.18.** *Let* $\langle \Delta_0, L_0 \rangle \to^* \langle \Delta_k, L_k \rangle$ *be a derivation starting from a valid initial state. Then,* $|\Delta_k| \leq |\mathtt{topterms}(\Delta_k) \setminus \mathcal{V}|^2\mathtt{maxarity}$.

*Proof.* We prove the lemma by induction on the length of the derivation, taking into account the strategy $(\mathtt{Shrink}^! \cdot \mathtt{ForcedDecompose})^!$. The lemma trivially holds for $\Delta_0$, since $|\Delta_0| = 1$. Note that the property of the lemma is preserved by the application of the four shrinking rules, namely `CycleOrClash`, `TwoNonVar`, `InvEq` and `NoSol`, since these rules do not increase the size of $\Delta$. Hence, it suffices to show that, if (i) a state $S = \langle \Delta_{k-1}, L_{k-1} \rangle$ satisfies the condition of the lemma, (ii) the rules `CycleOrClash`, `TwoNonVar`, `InvEq` and `NoSol` cannot be applied to $S$, and (iii) $S \to_{\mathtt{ForcedDecompose}} \langle \Delta_k, L_k \rangle$, then $|\Delta_k| \leq |\mathtt{subterms}(\mathtt{topterms}(\Delta_k)) \setminus \mathcal{V}|^2\mathtt{maxarity}$.

Consider the graph $G(\Delta_{k-1})$ (Definition V.5). Recall that the nodes $V$ of $G(\Delta_{k-1})$ is the set $\Delta^l_{k-1} \cup \mathtt{topterms}(\Delta_{k-1})$. First of all, note that $G$ must be acyclic due to the non-applicability of `InvEq` and hence it is a forest. Therefore, $\Delta_{k-1}$ is really a set; that is, $\Delta^l_{k-1}$ in Definition V.5 is the same as $\Delta_{k-1}$. We refer to the nodes in $V \cap \Delta_{k-1}$ as $m$-nodes, i.e. the nodes of $G$ that are multiequations and we refer to the nodes of $G$ that are non-variable terms as $f$-nodes.

Without loss of generality we assume that $G$ is a tree. We prove that the number $M$ of $m$-nodes is bounded by $T^2$, where $T$ is the number of $f$-nodes. (The following argument is the solution to the above puzzle.) Note that $T = |\mathtt{topterms}(\Delta_{k-1}) \setminus \mathcal{V}|$. We use induction on $M$. For base case, if $M = 1$, then $T \geq 3$ due to the non-applicability of `TwoNonVar`, and hence $M \leq T^2$ in this case.

For the inductive step, note that, again due to the non-applicability of `TwoNonVar`, $G$ must contain an $f$-node. Among all the $f$-nodes, pick one, say $v_f$, that has degree one: such an $f$-node must exist, since if every $f$-node has degree atleast 2, then the graph will have a cycle (note that every $m$-node has degree 3).

The $f$-node $v_f$ should have an edge to some $m$-node, say $v_m$. If we remove all outgoing edges from $v_m$, we should get atleast 3 disjoint trees (see illustration in Figure 5) that contain $f$-nodes (due to the non-applicability of `TwoNonVar`). One of those trees contains just one node – the $f$-node $v_f$. Assume that we get $l + l'$ other subtrees $G_1, \ldots, G_l, G_{l+1}, \ldots, G_{l+l'}$, and the first $l$ contain $f$-nodes. (The number $l$ is 2 in Figure 5

and $l'$ is 0.)

Let $M_i$ be the number of $m$-nodes in $G_i$, and let $T_i$ be the number of $f$-nodes in $G_i$, for all $i$. Note that $M_{l+1} = \cdots = M_{l+l'} = 0$ and $T_{l+1} = \cdots = T_{l+l'} = 0$, and that the trees $G_{l+i}$'s contain just one $v$-node each. Therefore, we have that $M = M_1 + \ldots M_l + 1$ and that $T = T_1 + \ldots T_l + 1$, where the last one is for the node $v_f$. Let us ignore the subtrees $G_{l+1}, \ldots, G_{l+l'}$.

By assumption, every $G_i$ contains an $f$-node, and fewer than $M$ $m$-nodes. We want to use the induction hypothesis, but before we can do that we need to make sure each $G_i$ satisfies our original constraint that every $m$-node has atleast 3 neighbours and that there is a path from each of those neighbours to a different $f$-node. Removing node $v_m$ may cause violation of this property. This happens if $v_m$ has an edge to a $v$-node in $G_i$. This is the case in the illustration in Figure 5 for the subgraphs that are encircled. We treat that $v$-node as an $f$-node and then apply induction hypothesis to get $M_i \leq (T_i + 1)^2$ for all $i$. Thus, we now have

$$\begin{aligned} M &= (\sum_{i=1}^{l} M_i) + 1 \\ &\leq \sum_{i=1}^{l} (T_i + 1)^2 + 1 \\ &= ((\sum_{i=1}^{l} T_i) + 1)^2 - \sum_{j \neq k} 2T_j T_k + l \\ &= T^2 - \sum_{j \neq k} 2T_j T_k + l \\ &\leq T^2 - (\sum_{j \neq k} 2) + l \\ &= T^2 - l(l-1) + l \quad \leq \quad T^2 \end{aligned}$$

For the last step in the above derivation, note that $l \geq 2$ and hence, $l \leq l(l-1)$ is always true.

Finally, note that the number of multiequations in $S$ is bounded by $|\texttt{topterms}(\Delta_{k-1}) \setminus \mathcal{Y}|^2$ and, since the application of `ForcedDecompose` increases the number of multiequations by at most a factor of `maxarity` we have that $|\Delta_k| \leq |\texttt{topterms}(\Delta_k) \setminus \mathcal{V}|^2 \texttt{maxarity}$. □

Every subproblem generated during a derivation will have polynomial size.

**Lemma V.19.** *Let $\langle \Delta_0, L_0 \rangle \rightarrow^* \langle \Delta_k, L_k \rangle$ be a derivation starting from a valid initial state. Then, for every multiequation $m \in |\Delta_k|$, $P(m)$ has polynomial size with respect to $k$.*

*Proof.* The lemma follows from Lemma A.1 (stated and proved in the appendix) and DefinitionV.2, and relies on the DAG representation for terms. Note that, thanks to the DAG representation, the terms in $L_k$ have polynomial size in $k$. □

Finally, our main result is the following, whose proof follows from the previous lemmas and can be found in the appendix.

**Theorem V.20.** *The 1-CU problem is solvable in polynomial time assuming a polynomial time oracle for 1-CU instances with at most two non-variable terms in the right hand-side of equations.*

## VI. One Context Unification Problems Solvable in Polynomial Time

The results in the previous section (Theorem V.20 and Lemmas IV.2 and IV.1) give us a reduction from the general 1-CU problem to the following restricted problem.

**Definition VI.1.** *An 1-CU instance $\mathcal{I}$ is called* reduced *if it is of the form*

$$\{F(u_i) \doteq x_i \mid i = 1, 2, \ldots\} \cup \{F(v_j) \doteq s \mid j = 1, 2, \ldots\} \cup \{F(w_k) \doteq t \mid k = 1, 2, \ldots\} \quad (1)$$

*where $s, t$ do not contain $F$; that is, the right hand-side of the equations have at most two non-variable terms.*

So far, we relied on an oracle to solve reduced instances. We will present special classes of 1-CU problems whose reduced instances can be solved in polynomial time. Certain reduced instances could have only one or two "non-trivial" equations. So, we first present results on solving 1-CU instances that have exactly one or two equations. These will help in solving more general reduced instances later.

### A. One Equation 1-CU Problem

We prove that a single equation 1-CU problem can be efficiently solved. If the equation is of the form $F(s) = C[F(t)]$ and $C$ is nonempty, then we can use Theorem IV.1 to solve it. Next, consider an equation of the form $F(C[F(s)]) = t$. It has the nice property that the hole position of any context that is a solution for $F$ can not be an extension of a *nonlinear* positions in $t$. A position $p$ is *nonlinear* in $t$ if there exists another position $q \neq p$ such that $t|_p = t|_q$. We also call $t|_p$ a nonlinear subterm of $t$.

**Lemma VI.2.** *Let $\mathcal{I}$ be a 1-CU instance consisting of one single equation of the form $F(C[F(s)]) \doteq t$ such that $F$ does not occur in $t$ (but $F$ can occur in $C$). Let $P = \{p \in \texttt{pos}(t) \mid t|_p = v$ and $v$ is a non-linear subterm of $t\}$. Then, $\forall p \in P : \texttt{hp}(F\sigma) \not\geq p$, for every solution $\sigma$ of $F(C[F(s)]) \doteq t$.*

*Proof.* Let $\sigma$ be a solution contradicting the conditions of the lemma, i.e. there is a term $v$ occurring at two distinct positions $p$ and $q$ of $t$ such that $\texttt{hp}(F\sigma) = p.p'$, form some $p'$. It follows that $F\sigma = t\sigma[\bullet]_{p.p'}$, $t\sigma[\bullet]_{p.p'}|_q = v\sigma$ and hence we have $v\sigma = C[F(s)]\sigma = C\sigma[F\sigma(s\sigma)] = C\sigma[t\sigma[\bullet]_{p.p'}[s\sigma]] = C\sigma[t\sigma[\bullet]_{p.p'}[s\sigma]]|_{\texttt{hp}(C).q} = v\sigma|_{\texttt{hp}(C).q} = v\sigma$, a contradiction. □

There are only a (linear number of) linear positions in a term. (In contrast, there can be exponentially many nonlinear positions in a term). It follows from Lemma VI.2 that for the equation $F(C[F(s)]) \doteq t$, we only need to test the (small number of) linear positions as possible hole positions. In fact,

we can enumerate a *complete* set of unifiers: a set of unifiers is *complete* if any other unifier (for the problem) is an instance of some unifier in the set. Here, a unifier is allowed to instantiate a context variable $F$ in terms of a new context variable $F'$.

**Lemma VI.3.** *Let $\mathcal{I}$ be a 1-CU instance consisting of one single equation of the form $F(C[F(s)]) \doteq t$ such that $F$ does not occur in $t$. Then, a complete set of unifiers $U$ of $\mathcal{I}$ of polynomial size can be computed in polynomial time. Any substitution $\sigma$ in $U$ satisfies one of the two conditions below:*

1) *Either $F\sigma = t[\bullet]_p$, with $p \in \mathtt{pos}(t)$,*
2) *Or $\sigma = \{F \mapsto t[F'(\bullet)]_q,\ x \mapsto F'(C[t[F'(s)]_q])\}$, where $x$ does not occur in $F(C[F(s)])$, $t|_q = x$, and $F'$ is a new context variable different from $F$.*

Proof of the above lemma can be found in the appendix. Using the special cases in Lemma VI.3 and Theorem IV.1, we can now prove that the special case when we have only one equation can be solved.

**Claim VI.1** (1-eqn)**.** *Let $\mathcal{I}$ be a 1-CU instance consisting of one single equation $F(s) \doteq t$, where $\mathtt{topsymbol}(t) \neq F$. Then, $\mathcal{I}$ can be solved in polynomial time.*

*Proof.* Note that the case where $F$ occurs in $t$ holds by Theorem IV.1. Note that the case where $F$ occurs in $s$ holds by the previous lemma. So, we are left with the case where $F$ does not occur in $s$ or $t$. In this case, it is easy to see that a unifier exists iff one exists where $\mathtt{hp}(F\sigma) \in \mathtt{pos}(t)$. To determine existence of solutions where $\mathtt{hp}(F\sigma) \in \mathtt{pos}(t)$, we just need to find a subterm of $t$ that unifies with $s$. Whereas the total number of subterms of $t$ might be exponential due to the DAG representation, the number of *distinct* subterms is polynomial. Hence, we can simply check, in polynomial time, all terms $v \in \mathtt{subterms}(t)$ that unify with $s$. □

### *B. Two Equation 1-CU Problem*

We show that 1-CU problems consisting of exactly two equations can be solved efficiently, but under a technical condition. To motivate the technical condition, note that the instance $\mathcal{I} = \{F(r_1) = z, F(r_2) = z\}$ can encode an *arbitrary* 1-CU instance $\mathcal{I}' = \{F(s_1) \doteq t_1, \ldots, F(s_n) \doteq t_n\}$, by having $r_1 = C[F(s_1), \ldots, F(s_n)]$ and $r_2 = C[t_1, \ldots, t_n]$. So, two equation case is as hard as an arbitrary number of equations. However, if we are interested in solutions $\sigma$ so that $\mathtt{hp}(F\sigma)$ is not below a position $p$ where both right-hand side terms have the same variable, then we can solve two equation problems.

First, consider two equations in which one has a nested $F$ on one side.

**Lemma VI.4.** *Let $\mathcal{I} = \{F(C[F(u)]) \doteq s, F(v) \doteq t\}$ be a 1-CU instance such that $s, t$ are non-variable terms not containing $F$. If we are only interested in solutions $\sigma$ such that*

$$\mathtt{not}(\exists p, x : (s|_p = t|_p = x \text{ and } \mathtt{hp}(F\sigma) > p)),$$

*then such a solution $\sigma$ of $\mathcal{I}$ can be found in polynomial time.*

Proof of the above lemma can be found in the appendix. Note that the ignored solutions make $s, t$ equal, but not all solutions that make $s, t$ equal are ignored.

Using the previous result, we can now solve the two equation case, but also when it is extended with some "variable definitions", under an extension of the same technical condition.

**Claim VI.2** (2-nonvar)**.** *Let $\mathcal{I}$ be a 1-CU instance of the form $\bigcup_{i=1}^{n}(\{F(u_i) \doteq x_i\}) \cup \{F(v_1) \doteq s, F(v_2) \doteq t\}$ such that the terms s,t, and $u_1, \ldots, u_n$ do not contain $F$, and $s, t, x_1, \ldots, x_n$ are pairwise different. If we are only interested in solutions $\sigma$ such that*

$$\begin{aligned}\mathtt{not}(\quad & s, t \text{ do not contain any } x_i, \text{ and} \\ & v_1 \text{ or } v_2 \text{ contains either } F \text{ or some } x_i, \text{ and} \\ & \exists p, x : (s|_p = t|_p = x \text{ and } \mathtt{hp}(F\sigma) > p)),\end{aligned}$$

*then such a solution $\sigma$ for $\mathcal{I}$ can be found in polynomial time.*

*Proof.* Note that, if some $x_i$ occurs in either $s$ or $t$, it has to occur properly, and the lemma follows from Lemma IV.2.

Hence assume that $s, t$ do not contain any $x_i$. We define the instance $\mathcal{I}'$ as the result of exhaustively replacing $x_i$ by $F(u_i)$ everywhere. If there is a cycle and this replacement can not be performed exhaustively, then $\mathcal{I}' = \bot$. Note that every solution of $\mathcal{I}'$ can be easily extended to be a solution of $\mathcal{I}$, that $\mathcal{I}'$ can be obtained from $\mathcal{I}$ in polynomial time, and that $||\mathcal{I}'||$ is polynomial w.r.t. $||\mathcal{I}||$. Moreover, $\mathcal{I}'$ is either of the form

1) $\bot$, or
2) $\{F(v_1) \doteq s, F(v_2) \doteq t\}$, where $v_1, v_2, s, t$ do not contain $F$ or
3) $\{F(v_1) \doteq s, F(v_2) \doteq t\}$, where $s, t$ do not contain $F$ and either $v_1$ or $v_2$ contains $F$.

In case 1 $\mathcal{I}'$ has no solutions. In case 2, $\mathcal{I}'$ is a 2-restricted 1-CU instance that we have shown can be efficiently solve in a companion paper [8]. Finally, case 3 follows from Lemma VI.4, while noting that the new technical condition maps to the condition in that lemma. □

### *C. Disjoint Variables and Constant Number of Equations*

Now that we have results for one and two equations, we next present sufficient conditions for polynomial time solvability of 1-CU problem. In each case, we will show that we can solve the corresponding reduced problems in polynomial time. Let us fix the following notation for the rest of the section.

$$\begin{aligned}\mathcal{I} &= \{F(s_1) \doteq t_1, \ldots, F(s_n) \doteq t_n\} \\ \mathcal{V}_1 &= \text{Set of all first-order variables in } s_1, \ldots, s_n \\ \mathcal{V}_2 &= \text{Set of all first-order variables in } t_1, \ldots, t_n \\ \mathcal{T}_1 &= \{u \mid u = F(u'),\ s_i|_p = u \text{ for some } i, p, u'\} \\ \mathcal{T}_2 &= \{u \mid u = F(u'),\ t_i|_p = u \text{ for some } i, p, u'\}\end{aligned}$$

Instance where $\mathcal{T}_2 \neq \emptyset$ can be solved in polynomial time using Theorem IV.1. Hence, the proofs below will implicitly assume $\mathcal{T}_2 = \emptyset$.

**Theorem VI.5.** *The class of 1-CU instances where $\mathcal{T}_1 = \emptyset$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ is solvable in polynomial time.*

*Proof.* (Sketch) Corresponding to any instance from this class, the reduced instances generated will also belong to this class. We solve reduced instances by unifying the left-hand sides corresponding to equal right-hand sides, and applying the unifier to the rest. Under the assumption that $\mathcal{T}_1 = \emptyset$, this unifier will be a first-order substitution. Since $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, the right-hand side terms do not change. The simplified reduced instance would be solved by either Theorem IV.1, or using Claim VI.1, or using Claim VI.2. See appendix for details. □

Theorem VI.5 subsumes the $\mathcal{V}_1 = \emptyset$ case, for which the best known algorithm is the NPtime procedure from [7]. In fact, we showed a family of examples from such a class in the introduction that would cause the algorithm from [7] to run in worst-case exponential time, whereas Theorem VI.5 would solve it in polynomial time.

Due to the technical condition in Claim VI.2, we could miss certain solutions that make two left-hand side terms (or equivalently, right-hand side terms) equal. But, if we have a *constant* number of equations, then we could systematically explore all possible combinations (of which left-hand sides are equal and which not) to get a complete procedure.

**Theorem VI.6.** *The class of 1-CU instances where $\mathcal{T}_1 = \emptyset$ and the cardinality $|\mathcal{I}|$ is assumed to be a constant $k$ (independent of the input problem size) is solvable in polynomial time.*

*Proof.* For each possible equivalence relation $\sim$ on $\mathcal{I}$, we do the following: If $F(s) \doteq t \ \sim \ F(s') \doteq t'$, we (first-order) unify $t, t'$ and $s, s'$ and apply the most-general unifier to all the equations in $\mathcal{I}$. Let $\mathcal{I}'$ be the new instance. Note that the number of equations in $\mathcal{I}'$ is equal to the number of equivalence classes induced by $\sim$. We then apply our inference rules on $\mathcal{I}'$. We need to show how to solve the reduced instances generated from $\mathcal{I}'$. We use the following algorithm for this purpose: If the reduced instance has two equations with identical right-hand sides, then we return "no solution" (for that reduced instance). If not, the reduced instance should have the form mentioned in Claim VI.2 and we solve it using Claim VI.2.

We claim that $\mathcal{I}$ has a solution iff for some $\sim$, the procedure above finds a solution. If the procedure above finds a solution, then clearly $\mathcal{I}$ has a solution. If $\mathcal{I}$ has a solution $\sigma$, define $\sim$ as follows: $F(s) \doteq t \ \sim \ F(s') \doteq t'$ if $s\sigma = s'\sigma$. When we use our procedure to solve instance $\mathcal{I}'$ generated from $\mathcal{I}$ using $\sim$, we are guaranteed to find a solution. This is because, while Claim VI.2 can miss solutions, it never misses a solution that makes all left-hand side terms (in $\mathcal{I}'$) different. □

We can extend the previous result to allow $\mathcal{T}_1 \neq \emptyset$.

**Corollary VI.7.** *The class of 1-CU instances where the cardinality $|\mathcal{T}_1|+|\mathcal{I}|$ is assumed to be a constant $k$ (independent of the input problem size) is solvable in polynomial time.*

*Proof.* If $\mathcal{T}_1 \neq \emptyset$, then we introduce new variables and get an instance where $\mathcal{T}_1 = \emptyset$. Specifically, if $u \in \mathcal{T}_1$, then we add the equation $u \doteq x$ to the instance (where $x$ is a new variable), and replace $u$ by $x$ everywhere else. Applying this repeatedly, we get an instance for which $\mathcal{T}_1 = \emptyset$, and now we can use Theorem VI.6. □

### *D. Left- and Right-Ground 1-CU Problems*

Right-ground 1-CU instances can be efficiently solved.

**Theorem VI.8.** *The class of 1-CU instances where $\mathcal{V}_2 = \emptyset$ is solvable in polynomial time.*

*Proof.* If the instance has 2 equations, we solve it using Claim VI.2. Note that we do not miss any solutions due to the technical condition. If the instance has 3 or more equations, we use our inference rules. The procedure will not generate any reduced instances to solve since all ($\geq 3$) right-hand side terms are always ground (in any generated subproblem). □

The result in Theorem VI.8 was already known [6], but now we can a new proof using our procedure. We can generalize Theorem VI.8 to a class that does not require all right-hand sides terms to be ground, but just two of them.

**Theorem VI.9.** *Let $\mathcal{I}$ be a 1-CU instance of the form $\mathcal{I}' \cup \{F(s_1) \doteq s, F(s_2) \doteq t\}$, where $s, t$ are distinct ground terms. Then $\mathcal{I}$ can be solved in polynomial time.*

*Proof.* Clearly, for every solution $\sigma$, $\mathtt{hp}(F\sigma) \in \mathtt{pos}(s) \cap \mathtt{pos}(t)$. Hence, $l = |\mathtt{hp}(F\sigma)|$ is polynomial even when $s$ and $t$ are DAGs. We now argue that, once $l$ is fixed, there is only *one* choice for $\mathtt{hp}(F\sigma)$, and thus also $F\sigma$. If $l = 0$ the claim holds trivially. hence assume $l > 0$ and, without loss of generality, let $s = f(s_1, s_2)$ and $t = f(t_1, t_2)$. Note that, since $s \neq t$, $\exists i \in \{1, 2\} : t_i \neq s_i$ holds. Also note that, if $\forall i \in \{1, 2\} : t_i \neq s_i$ holds then there is no solution of length $l$. In the remaining case either $s_1 = t_1$ or $s_2 = t_2$, say $s_2 = t_2$. Then $hp(F\sigma) = 1.\mathtt{hp}(F')$, $F\sigma = t[F'(\bullet)]_1$, and, since $|\mathtt{hp}(F')| < l$ the claim holds by induction hypothesis.

Since there are only polynomially many choices for $\mathtt{hp}(F\sigma)$, we can enumerate them all and solve $\mathcal{I}$ using polynomially many calls to a first-order unification procedure. □

Left-ground 1-CU instances can also be efficiently solved.

**Theorem VI.10.** *The class of 1-CU instances where $\mathcal{V}_1 = \mathcal{T}_1 = \emptyset$ is solvable in polynomial time.*

*Proof.* We apply our inference rules and use the following algorithm to solve the generated reduced instances: If the reduced instance has two equations with identical right-hand sides, then we return "no solution" (for the reduced instance) if the left-hand sides are not identical, and we delete one equation from $\mathcal{I}$ if the left-hand sides are identical. Let $\mathcal{I}'$ be the new reduced instance. Let $F(u) \doteq x$ be an equation in $\mathcal{I}'$. If $x$ occurs in any other right-hand side term in $\mathcal{I}'$, then we use Theorem IV.1. Since $\mathcal{V}_1 = \emptyset$, $x$ can not occur in left-hand side terms. So, we can remove $F(u) \doteq x$ from $\mathcal{I}'$. Hence, finally $\mathcal{I}'$ will have at most two equations. If $\mathcal{I}'$ has zero or one equations, then we are done by Claim VI.1. So, assume $\mathcal{I}'$ has exactly two equations, say $F(v) \doteq s$ and $F(w) \doteq t$. If

$v \neq w$, we can use Claim VI.2 (we will not miss any solutions since $v$ and $w$ are ground and therefore they will be different in all solutions.) If $v = w$, then we unify $s$ and $t$ and solve the resulting one equation. □

Finally, we generalize Theorem VI.10 to remove the requirement $\mathcal{T}_1 = \emptyset$, and consider instances where only $\mathcal{V}_1 = \emptyset$. Note that the context matching problem matching with a constant number of context variable was solved in polynomial time in [6], but the class $\mathcal{V}_1 = \emptyset$ falls out of the class solved in [6].

**Theorem VI.11.** *The class of 1-CU instances where $\mathcal{V}_1 = \emptyset$ is solvable in polynomial time.*

*Proof.* We follow along the lines of the proof of Theorem VI.10. We now use the following algorithm to solve any generated reduced instance, say of form given in Equation 1: If the reduced instance $\mathcal{I}'$ has two equations $F(u) \doteq r$ and $F(u') \doteq r$ with identical right-hand sides, then we use (first-order) decomposition rule on $u \doteq u'$ exhaustively to get a set of equations, where each equation is of the form $F(u_1) \doteq u_2$, where $F(u_1) \in \mathcal{T}_1$ and $u_2$ is a first-order ground term whose top symbol is not $F$, but $u_2$ possibly contains $F$. We add these equations to the reduced instance and remove one of the original equations, say $F(u) \doteq r$, from $\mathcal{I}'$. If $u_2$ contains $F$, then we can solve $\mathcal{I}'$ using Theorem IV.1. Hence, assume the first-order ground term $u_2$ does not contain $F$.

Let $\mathcal{I}''$ denote the reduced instance we get after we have processed all repeated right-hand side equations as above. Let $F(u) \doteq x$ be an equation in $\mathcal{I}''$. If $x$ occurs in any other right-hand side term in $\mathcal{I}''$, then we use Theorem IV.1. Since $\mathcal{V}_1 = \emptyset$, $x$ can not occur in left-hand side terms. So, we can remove $F(u) \doteq x$ from $\mathcal{I}''$ without changing its solvability. Hence, finally $\mathcal{I}''$ will have the form:

$$F(v) \doteq s, F(w) \doteq t, F(u_1) \doteq u'_1, \ldots, F(u_k) \doteq u'_k$$

where $s, t$ are the non-variable terms in the original reduced instance, and the last $k$ equations are generated from unification of left-hand sides. The terms $u'_1, \ldots, u'_k$ are all ground. We can assume all $u'_i$ are distinct, otherwise we could just repeat the above process.

If $k \geq 2$, we are done by Theorem VI.9.

If $k = 0$, then we solve using Claim VI.2, but to overcome the incompleteness there, we additionally solve the instance $\mathcal{I}'''$ obtained by unifying $s$ and $t$, and then adding equations obtained by decomposing $v \doteq w$. The instance $\mathcal{I}'''$ has at most one non-ground right-hand side. If it has two or more distinct ground right-hand sides, then again we can use Theorem VI.9. If it has exactly one ground right-hand side, and one non-ground, then we Claim VI.2 again, and this time, the instance we need to solve to overcome the incompleteness there, will have only ground right-hand sides (for which we can use either Theorem VI.9 or Claim VI.1).

If $k = 1$, then we have 3 equations, where one has a ground right-hand side. We use our inference system on these three equations to obtain new reduced instances. Each new reduced instance has 3 equations: one has a variable on the right, one has a ground term, and the third can have an arbitrary first-order term. We can again remove the equation with a variable on the right and get instances like in case $k = 0$ above. □

### E. The General 1-CU Problem

We can actually use our procedure to solve the general 1-CU problem in stages as follows: (a) first we reduce an instance $\mathcal{I}$ to polynomially many reduced instances $\mathcal{I}_1, \ldots, \mathcal{I}_k$, (b) for each reduced instance $\mathcal{I}_j$, we unify left-hand sides that have equal right-hand sides, apply the substitution and obtain new reduced instances $\mathcal{I}'_1, \ldots, \mathcal{I}'_k$, and (c) finally we apply our procedure recursively on each new reduced instance. We *conjecture* that this procedure yields a polynomial time algorithm for the general 1-CU problem.

## VII. Conclusion

We presented an inference system for solving the one context unification problem. We proved that the inference system yields a polynomial time algorithm for several classes of one context unification problems. The inference system itself has many interesting features: the proof search continues along one main branch, while the side branches are immediately terminated using polynomial time procedures. The main branch itself can generate a large number of subproblems, but their number is bounded using an interesting graph argument, which could be of independent interest.

## References


[1] Franz Baader and Wayne Snyder. Unification theory. In John Alan Robinson and Andrei Voronkov, editors, *Handbook of Automated Reasoning (in 2 volumes)*, pages 445–532. Elsevier and MIT Press, 2001.

[2] Carles Creus, Adrià Gascón, and Guillem Godoy. One-context Unification with STG-Compressed Terms is in NP. In *RTA*, pages 149–164, 2012.

[3] Robert Dabrowski and Wojciech Plandowski. Solving two-variable word equations (extended abstract). In *Automata, Languages and Programming: 31st International Colloquium, ICALP 2004, Turku, Finland, July 12-16, 2004. Proceedings*, pages 408–419, 2004.

[4] Robert Dabrowski and Wojciech Plandowski. On word equations in one variable. *Algorithmica*, 60(4):819–828, 2011.

[5] Katrin Erk and Joachim Niehren. Dominance constraints in stratified context unification. *Inf. Process. Lett.*, 101(4):141–147, 2007.

[6] Adrià Gascón, Guillem Godoy, and Manfred Schmidt-Schauß. Context Matching for Compressed Terms. In *LICS*, pages 93–102, 2008.

[7] Adrià Gascón, Guillem Godoy, Manfred Schmidt-Schauß, and Ashish Tiwari. Context Unification with One Context Variable. *J. Symb. Comput.*, 45(2):173–193, 2010.

[8] Adrià Gascón, Manfred Schmidt-Schauss, and Ashish Tiwari. One context unification and fault-tolerant unification, 2015. http://www.csl.sri.com/users/tiwari/1cu-2r.pdf.

[9] Warren D. Goldfarb. The Undecidability of the Second-Order Unification Problem. *Theor. Comput. Sci.*, 13:225–230, 1981.

[10] Sumit Gulwani and Ashish Tiwari. Computing procedure summaries for interprocedural analysis. In *16th European Symposium on Programming Languages and Systems, ESOP 2007, part of the Joint European Conferences on Theory and Practics of Software, ETAPS 2007, Braga, Portugal, March 24 - April 1, 2007*, pages 253–267, 2007.

[11] Artur Jeż. One-variable word equations in linear time. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II*, pages 324–335, 2013.

[12] Artur Jeż. Context unification is in PSPACE. In *Proc. ICALP 2014, Part II*, volume 8573 of *Lecture Notes in Computer Science*, pages 244–255. Springer, 2014.

[13] Jordi Levy, Joachim Niehren, and Mateu Villaret. Well-Nested Context Unification. In *20th International Conference on Automated Deduction, Tallinn, Estonia, July 22-27, 2005, Proceedings*, pages 149–163, 2005.
[14] Jordi Levy, Manfred Schmidt-Schauß, and Mateu Villaret. On the complexity of bounded second-order unification and stratified context unification. *Logic Journal of the IGPL*, 19(6):763–789, 2011.
[15] G. S. Makanin. On the decidability of the theory of free groups (in russian). In *Fundamentals of Computation Theory, FCT '85, Cottbus, GDR, September 9-13, 1985*, pages 279–284, 1985.
[16] Alberto Martelli and Ugo Montanari. An efficient unification algorithm. *ACM Trans. Program. Lang. Syst.*, 4(2):258–282, 1982.
[17] S. Eyono Obono, Pavel Goralcik, and M. N. Maksimenko. Efficient solving of the word equations in one variable. In *Mathematical Foundations of Computer Science 1994, 19th International Symposium, MFCS'94, Kosice, Slovakia, August 22 - 26, 1994, Proceedings*, pages 336–341, 1994.
[18] Mike Paterson and Mark N. Wegman. Linear unification. *J. Comput. Syst. Sci.*, 16(2):158–167, 1978.
[19] Wojciech Plandowski. Satisfiability of word equations with constants is in PSPACE. *J. ACM*, 51(3):483–496, 2004.
[20] John Alan Robinson. A Machine-Oriented Logic Based on the Resolution Principle. *J. ACM*, 12(1):23–41, 1965.
[21] Manfred Schmidt-Schauß and Jürgen Stuber. The complexity of linear and stratified context matching problems. *Theory Comput. Syst.*, 37(6):717–740, 2004.

## APPENDIX

**Lemma A.1.** *Let $\langle \Delta_0, L_0 \rangle \rightarrow^* \langle \Delta_k, L_k \rangle$ be a derivation starting from some valid initial state (corresponding to some 1-CU instance). Let $T$ be* $\texttt{subterms}(\texttt{topterms}(\Delta_k) \setminus \mathcal{Y})$*. The possible substitutions applied to the state $\langle \Delta_k, L_k \rangle$ by the inference rules are of the forms*

1) *$x \mapsto t$, where $x \in \mathcal{X}$, $t \in T$,*
2) *$y \mapsto t$, where $y \in \mathcal{Y}$ and $t \in T$,*
3) *$y_1 \mapsto y_2$, where $y_1, y_2 \in \mathcal{Y}$, and*
4) *$x \mapsto f(y_1, \ldots, y_m)$, where $x \in \mathcal{X}$ and $y_1, \ldots, y_m \in \mathcal{Y}$.*

*Moreover,* $\texttt{subterms}(\texttt{topterms}(\Delta_k) \setminus \mathcal{Y}) \cap \mathcal{Y} = \emptyset$*, i.e. the variables in $\mathcal{Y}$ only occur at the top in the terms in* $\texttt{topterms}(\Delta_k)$.

*Proof.* The lemma follows by induction on the length of the derivation, distinguishing cases according to the last applied rule. The only non-trivial case is the `ForcedDecompose` rule, where variables from $\mathcal{Y}$ are introduced. This case follows from the maximality of $\Gamma$ in the rule application, and the fact $\texttt{ForcedDecompose}(\langle \Delta, L \rangle, \Gamma, \mathcal{Y})$ decomposes every instantiated term, under the assumption that variables from $\mathcal{Y}$ are always instantiated in terms of variables from $\mathcal{X}$. □

**Lemma A.2** (Lemma V.15)**.** *Let $S = \langle \Delta, L \rangle$ be a state of our procedure such that no rule can be applied to $S$. Then, $\Delta = \emptyset$.*

*Proof.* We prove by contradiction. Assume that $\Delta$ is non-empty and no rule can be applied to $S$. In particular, the `ForcedDecompose` rule is not applicable. This can happen if (i) either the (strict) subterm relation on $\texttt{topterms}(\Delta)$ is cyclic, which caused failure to find an appropriate $\Gamma$,
(ii) or we could find an $\Gamma$, but $|\texttt{topsymbols}(\Gamma)| \neq 1$.

First assume that (i) holds. Then, there must be multiequations $m_0, \ldots, m_k$ in $\Delta$, terms $s_0, \ldots, s_k$ and contexts $C_0, \ldots, C_k$ satisfying $s_i \in m_i$ and $C_i[s_i] \in m_{(i+1)\%(k+1)}$, for all $i \in \{0, \ldots, k\}$. Note that $\texttt{mgu}(\{m_1, \ldots, m_k\}) = \bot$. This implies that $\Delta = \{m_0, \ldots, m_k\}$, since otherwise `NoSol` would be applicable. Moreover, for all $i$, $\texttt{mgu}(\{m_0, \ldots, m_k\} \setminus \{m_i\}) \neq \bot$, again because `NoSol` is not applicable. Hence, by Definition V.2, every 1-CU instance $P(m_i)$ contains two equations of the form $F(u) \doteq s'_i, F(v) \doteq C'_i[s'_i]$ and hence in particular $P(m_0)$ satisfies the conditions of Lemma IV.2. This implies that the `CycleOrClash` rule is applicable, a contradiction.

Now consider the case when (ii) holds. First assume that $|\texttt{topsymbols}(\Gamma)| = 0$. In this case, for every multiequation $m$ in $\Gamma$ contains only variables and these variables do not occur as subterms in other multiequations in $\Delta$. Hence, these variables are not instantiated by $\texttt{cmgu}(m)$ and hence $\texttt{topnonvars}(m\texttt{cmgu}(m)) = 0$. Hence, rule `TwoNonVar` is applicable in this case, a contradiction.

Finally, assume that $|\texttt{topsymbols}(\Gamma)| > 1$. Without loss of generality, among all possible choices for $\Gamma$, pick one that is of smallest cardinality. For such a $\Gamma$, if $|\texttt{topsymbols}(\Gamma)| > 1$ and $m \in \Gamma$, then it is also the case that $|\texttt{topsymbols}(m\texttt{cmgu}(m))| > 1$. Thus, there is multiequation $m \in \Delta$ such that $P(m)$ satisfies the conditions of Lemma IV.1. Hence, `CycleOrClash` is applicable, which is a contradiction. □

**Theorem A.3** (Theorem V.20)**.** *The 1-CU problem is solvable in polynomial time assuming a polynomial time oracle for 1-CU instances with at most two non-variable terms in the right hand-side of equations.*

*Proof.* Let $\mathcal{I}$ be a 1-CU instance and let $\langle \Delta_0, L_0 \rangle \rightarrow^* \langle \Delta_k, L_k \rangle$ be its corresponding derivation in the our algorithm. By Lemma V.17, $|\texttt{subterms}(\texttt{topterms}(\Delta_i) \setminus \mathcal{Y})| \leq |\texttt{subterms}(\texttt{topterms}(\Delta_0) \setminus \mathcal{Y})| \leq ||\mathcal{I}||$ for all $i$. In other words, the number of different non-variable subterms in the $\Delta_i$s does not increase. By Lemma V.18, $|\Delta_i| \leq |\texttt{topterms}(\Delta_i) \setminus \mathcal{V}|\texttt{maxarity}$, for all $i$, and hence the size of the $\Delta_i$s is always bounded by $|\texttt{subterms}(\texttt{topterms}(\Delta_0) \setminus \mathcal{Y})|\texttt{maxarity} \leq ||\mathcal{I}||\texttt{maxarity}$.

Note that an application of a shrinking rule either reduces the number of multiequations in $\Delta_i$ or unifies two terms in $\texttt{topterms}(\Delta_i)$, for $i \in \{0, \ldots, k\}$. It follows that a sequence of applications of shrinking rules has length at most $n|\texttt{subterms}(\texttt{topterms}(\Delta_0) \setminus \mathcal{Y})|\texttt{maxarity} \leq |\mathcal{Y}|$, where $n$ is the size of the multiequations in the $\Delta_i$s, i.e. the number of equations in $\mathcal{I}$.

Finally, by Lemma V.17, every application of `ForcedDecompose` reduces $|\texttt{subterms}(\texttt{topterms}(\Delta_i)) \setminus \mathcal{V}|$ and hence every derivation contain at most $||\mathcal{I}||$ application of this rule, which gives a quadratic bound $||\mathcal{I}||^2\texttt{maxarity}$ for the length of any derivation. To conclude, note the rules can be checked for applicability, applied, and the corresponding spanned problems can be generated in polynomial time and have polynomial size by Lemma V.19. □

**Lemma A.4** (Lemma VI.3)**.** *Let $\mathcal{I}$ be a 1-CU instance consisting of one single equation of the form $F(C[F(s)]) \doteq t$ such that $F$ does not occur in $t$. Then, a complete set of unifiers $U$ of $\mathcal{I}$ of polynomial size can be computed in polynomial time.*

*Any substitution $\sigma$ in $U$ satisfies one of the two conditions below:*

1) *Either $F\sigma = t[\bullet]_p$, with $p \in \mathtt{pos}(t)$,*
2) *Or $\sigma = \{F \mapsto t[F'(\bullet)]_q,\ x \mapsto F'(C[t[F'(s)]_q])\}$, where $x$ does not occur in $F(C[F(s)])$, $t|_q = x$, and $F'$ is a new context variable different from $F$.*

*Proof.* We distinguish two cases. First consider solutions $\sigma$ satisfying that $\mathtt{hp}(F\sigma) \in \mathtt{pos}(t)$. By Lemma VI.2, for each of such solutions, $\mathtt{hp}(F\sigma)$ must be in the set $Q = \{p \in \mathtt{pos}(t) \mid t|_p = v$ and $v$ is a linear subterm of $t\}$. Note that $\{F \to t[\bullet]_p\} \leq \sigma$. Since $|Q|$ is polynomial w.r.t. $|t|$ even in the DAG representation, then $U$ has a polynomial number of solutions of this form. Now consider solutions $\sigma$ such that $\mathtt{hp}(F\sigma) \notin \mathtt{pos}(t)$. Let $p = p_1.p_2$, where $p_1$ is the longest prefix of $p$ defined in $t$. Note that $t|_{p_1}$ must be a variable $x$ linear in $t$ by Lemma VI.2. Moreover, since $\sigma$ satisfies $x\sigma|_{p_2} = C[F(s)]\sigma$, $x$ does not occur in $C[F(s)]$. Hence, $\sigma$ is of the form $\{F \mapsto t[D]_{p_1}, x \mapsto DC[t[D(s)]_{p_1}]\}$, for an arbitraty context $D$ such that $\mathtt{hp}(D) = p_2$. Hence, all solutions $\sigma$ such that $\mathtt{hp}(F\sigma) = q.q'$ and $q.q' \notin \mathtt{pos}(t)$, with $q \in Q$, can be represented by a substitution $\theta = \{F \mapsto t[[F'(\bullet)]]_q, x \mapsto F'(C[t[F'(s)]_q])\}$, where $t|_q = x$, $F'$ is a new context variable different from $F$, since $\theta \leq \sigma$ holds. □

**Lemma A.5** (Lemma VI.4)**.** *Let $\mathcal{I} = \{F(C[F(u)]) \doteq s, F(v) \doteq t\}$ be a 1-CU instance such that $s, t$ are non-variable terms not containing $F$. If we are only interested in solutions $\sigma$ such that*

$$\mathtt{not}(\exists p, x : (s|_p = t|_p = x \ \mathtt{and}\ \mathtt{hp}(F\sigma) > p)),$$

*then such a solution $\sigma$ of $\mathcal{I}$ can be found in polynomial time.*

*Proof.* If either $s$ or $t$ is a constant, the lemma is straighforward. Hence, assume that $s$ and $t$ are both not constants. By Lemma VI.3 we can compute, in polynomial time, a complete set of unifiers $U = \theta_1, \ldots, \theta_k$ of the single equation $F(C[F(u)]) \doteq s$ of polynomial size. Moreover, every substitution $\theta \in U$ satisfies one of the two conditions below:

1) $F\sigma = s[\bullet]_p$, with $p \in \mathtt{pos}(s)$, or
2) $\sigma = \{F \mapsto s[F'(\bullet)]_q, x \mapsto F'(C[s[F'(u)]_q])\}$, where $x$ does not occur in $F(C[F(u)])$, $s|_q = x$, and $F'$ is a new context variable different from $F$.

Hence, to obtain a polynomial time algorithm, it is enough to check if some substitution in $U$ can be extended to solve also the equation $F(v) \doteq t$, and thus $\mathcal{I}$. Consider the two cases of a substitution $\theta \in U$.

(1) If $\theta$ is such that $F\sigma = s[\bullet]_p$, then the check can clearly be done in polynomial time, since $F(v)\theta \doteq t\theta$ is a first-order unification instance of polynomial size thanks to the DAG representation.

(2) Otherwise, $\theta$ is of the form $\{F \mapsto s[F'(\bullet)]_q, x \mapsto F'(C[s[F'(u)]_q])\}$, where $s_q = x$, we distinguish cases depending on whether $x$ occurs in $t$, and if so, where.

(i) First assume that $x$ occurs in $t$ at a position $p$ such that either $p < q$ or $p > q$. Since we are looking for solution $\sigma$ where $\mathtt{hp}(F\sigma) \geq q$, these cases can be rewritten into a form that is covered by Theorem IV.1. Note that since $p \neq q$, $C$ in Theorem IV.1 is nonempty.

(ii) Assume that, for some $p$, $t|_p = x$ and $p$ is disjoint with $q$. In this case, every solution $\sigma$ that is an extension of $\theta$ satisfies that $|F\sigma| > |F\sigma|_q\sigma| = |x\sigma| = |F(C[F(u)])\sigma| > |F\sigma|$, a contradiction. Hence we know that if $x$ occurs in $t$ at a position disjoint with $q$ there are no solutions that are extensions of $\theta$ and thus there is no need to test $\theta$.

(iii) Consider now the case where $t|_q = x$. We are explicitly not interested in these solutions.

(iv) Finally consider the case where $x$ does not occur in $t$. Then we have $(F(v)\theta \doteq t\theta) = (F(v)\theta \doteq t) = s[F'(v\theta)]_q \doteq t$. Note that, since $s$ does not contain $F$, we can use a few (first-order unification) `Decompose` steps to get a *single* equation $F'(v') \doteq t'$ that can be solved by Claim VI.1. □

**Theorem A.6** (Theorem VI.5)**.** *The class of 1-CU instances where $\mathcal{T}_1 = \emptyset$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ is solvable in polynomial time.*

*Proof.* Corresponding to any instance from this class, the reduced instances generated will also belong to this class. Consider a reduced instance of form shown in Equation 1. We unify the left-hand sides corresponding to equal right-hand sides, and apply the unifier to the rest. Under the assumption that $\mathcal{T}_1 = \emptyset$, this unifier will be a first-order substitution. Since $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, the right-hand side terms do not change. Thus, the simplified reduced instance would be of the form $\{Fu_1 \doteq x_1, \ldots, Fu_m \doteq x_m, Fv \doteq w, Fv' \doteq w'\}$, where $x_i$'s are all different. If any $x_i$ occurs in $w, w'$, we are done by Theorem IV.1. If not, we can remove equations with $x_i$ on right-hand side, and just solve $\{Fv \doteq w, Fv' \doteq w'\}$ in two steps: first, we unify $w, w'$ and $v, v'$ and get one equation and solve it using Claim VI.1, and if we do not find a solution to that one equation, then we find a solution for the two equations using Claim VI.2. Note that (a) the first step (one equation) guarantees that we do not miss any solutions that are missed due to the technical condition in Claim VI.2, and (b) it was possible to do so because $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ and unifiers of terms on one side do not instantiate the other side. □